

Using SAFMEDS to assist language learners to acquire second language vocabulary

Beverley, Michael; Hughes, J. Carl; Hastings, Richard

European Journal of Behavior Analysis

DOI:

[10.1080/15021149.2016.1247577](https://doi.org/10.1080/15021149.2016.1247577)

Published: 01/11/2016

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](https://doi.org/10.1080/15021149.2016.1247577)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Beverley, M., Hughes, J. C., & Hastings, R. (2016). Using SAFMEDS to assist language learners to acquire second language vocabulary. *European Journal of Behavior Analysis*, 17(2), 131-141. <https://doi.org/10.1080/15021149.2016.1247577>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Using SAFMEDS to assist language learners to acquire second
language vocabulary

Michael Beverley¹, J. Carl Hughes¹, and Richard P. Hastings²

¹School of Psychology, Bangor University, Bangor, Gwynedd, Wales.

²Centre for Educational Development, Appraisal and Research (CEDAR)

University of Warwick, Coventry.

Suggested Running Head: Using SAFMEDS to assist second language learners

Submitting Author: Michael Beverley

School of Psychology

Bangor University

Brigantia Building

Gwynedd, Wales, LL57 2AS

Telephone: 01248 3832467

Email: m.beverley@bangor.ac.uk

Abstract

SAFMEDS (*Say All Fast Minute Every Day Shuffled*) cards, in conjunction with graphing learner's progress, were used as an intervention to teach Welsh vocabulary to second language learners in four year-7 classes (12-year-olds) in an English medium secondary school in Wales. A further class of children of the same age served as a Waiting List Control (WLC). From the curriculum, 200 words were selected by teachers to be studied during the school term. Children in all classes were given a pre-test to establish the number of words that each child already knew at baseline. Children in the four intervention classes were each given packs of Welsh-English SAFMEDS and required to pair up with another child. Both children were instructed to conduct a practice run followed by a 1-minute timing at the start of each 50-minute lesson (3 lessons per week). Children then recorded their correct and incorrect responses on data sheets and graphs. The procedure took less than 5 minutes, and the teacher then continued with the scheduled lesson. The WLC class received their normal Welsh classes. The intervention spanned four weeks of the term and was followed with a post-test of the target words for all children. During the post-test, the intervention children wrote significantly more correct Welsh vocabulary words than the control children, $p < .001$, $d = 1.54$. The study demonstrates that SAFMEDS are an efficient and effective method to enhance vocabulary learning with brief exposure within the context of standard lessons.

Keywords: SAFMEDS, Reliable Change Index (RCI), Numbers Needed to Treat (NNT), basic skills, fluency-building procedures, second language vocabulary learning.

For any language, vocabulary is a key component when learning to speak, read, and write. Although vocabulary acquisition forms only one element of a comprehensive language learning programme, it nevertheless plays a central role in language acquisition and is of particular importance to beginner language learners (Coady & Huckin, 1997). There has been an increased interest in vocabulary learning since the 1970s (Carter, 1987) and current research supports the position that second-language vocabulary can be learned through four main methods: Direct teaching—teacher explanation or peer teaching; Direct learning—using dictionary or word cards; Incidental learning—guessing from context; or Planned encounters—graded reading or vocabulary exercises (Nation, 2001).

Although learning a second-language through incidental reading has been said to be one of the primary means of the continued development of a learner's vocabulary, according to Huckin and Coady (1999), any learner must already have a sight recognition vocabulary of approximately 3,000 words for this strategy to be effective. Other research also supports the effectiveness of deliberate learning of vocabulary, especially at the beginning stages of learning (Elgort, 2011; Kang & Golden, 1994), and that learners need to take responsibility for their own learning for it to be successful (Nation, 2008).

A typical passage of text contains a number of different types of vocabulary that fall into four distinct categories (i.e., high-frequency, low-frequency, academic, and technical words; (Nation, 2001). Therefore, for learners of second language vocabulary, it may be crucial to include direct teaching of these high frequency words early in the learning journey so students can begin to negotiate reading and listening to the new language with comprehension. These strategies may also be as equally applied to the general learning of an individual's first-language as to any second-language learning (August, Carlo, Dressler, & Snow, 2005; Mueller Gathercole, Mon Thomas, & Hughes, 2008; NICHD, 2000).

Various approaches have been used to directly teach high-frequency vocabulary including using mobile phone applications (Lu, 2008); language laboratories and other audio-visual media (Vanderplank, 2009); and keyword, mnemonic techniques, and imagery (Beaton, Gruneberg, & Ellis, 1995). Supporting evidence for these techniques varies and suggests that keyword vocabulary learners do not maintain their learning over time (Wang & Thomas, 1995). For learners to retain second-language vocabulary over extended periods without on-going practice (maintenance strategy), it has been suggested that there must also be a *fluency development* aspect to any instruction, whereby students are enabled to become more proficient with words they already know (Nation, 2008). This requires that educators design the curriculum to allow the learner to repeatedly cover the most useful language items, and practice these items to a fluent level of performance (Binder, 1996; Binder, Haughton, & Bateman, 2002; Binder, Haughton, & Van Eyk, 1995; Nation, 2008).

The focus for the present study was on the learning of language in a mainstream school in the context of second language learning. Precision Teaching (PT) and fluency based instruction tactics and strategies have previously been used successfully to teach academic tasks to many individuals and across diverse curriculum subjects (Binder et al., 2002; Bloom, 1986; Hughes, Beverley, & Whitehead, 2007; Kubina & Yurich, 2012). Although PT methods have typically been used with small numbers of participants we could find no examples of the use of these methods for learning a second language and we wanted to evaluate the use of these procedures within whole classes in the context of standard lessons (Beverley, Hughes, & Hastings, 2009; Roberts & Norwich, 2010). It has been shown that too often any teaching methods that are introduced in a school setting are not continued once the researchers or other implementing bodies leave the organisation (Georgiades & Phillimore, 1975; Roberts & Hampton, 2008), and especially if the organisations receive no further training or coaching in those methods (Johnson & Street, 2004).

The aim of the research from the school's perspective was to intervene to buck the current trend with existing practices. Teachers had noted that children from these classes tended to do poorly on learning new Welsh vocabulary words, even though this was the main aim of the class. Our remit from the school was that throughout the entire study we would enable the teachers to be in control and ultimately responsible for the introduction and running of the intervention within each classroom. This, in turn, would allow teachers to learn about these methods through active engagement in their normal classroom settings. Therefore, whilst the research team designed the intervention and evaluation methodology, the teachers ran the intervention at the beginning of each class, throughout the entire study.

We decided to use one of the tools often used in PT methodologies—SAFMEDS. SAFMEDS (*Say All Fast Minute Every Day Shuffled*) are a practice and assessment procedure developed to help students learn and build fluency on key facts (Graf & Lindsley, 2002). SAFMEDS are typically used to help students become fluent in definitions and basic concepts, and were therefore ideally suited to practicing vocabulary of two languages. Further details of SAFMEDS usage have been described in earlier publications (Claypool-Frey, 2009; Graf & Lindsley, 2002; Vieitez, 2003).

We designed a brief intervention that would allow children many opportunities for active responding (Binder, 1996; Fredrick & Hummel, 2004; Heward, 1994; Johnson & Layng, 1996). Therefore, the intervention we implemented was a simple, direct way of teaching. It was a short, sharp, focused, fluency-based method, which was not resource intensive but was easy to teach to the children and easy for teachers to implement.

The purpose of the study was to compare the performance of children who received the intervention (fluency-based practice using SAFMEDS cards) with the children who did not receive the intervention. We wished to measure whether children who received this brief intervention would be able to correctly write more Welsh vocabulary words (from a

randomly sorted list of their English equivalents) at post-test than the children who did not receive the intervention.

Method

Participants and Setting

The study took place with five classes of school children within one mainstream secondary school in North Wales. There were 95 children in the sample, with 50 males and 45 females. All children were between 11-12 years old. One class from the five selected was randomly allocated to be the Waiting List Control group (WLC; $n = 16$), the other four classes ($n = 79$) were allocated to receive the intervention: Intervention Groups 1 (I1; $n = 25$), 2 (I2; $n = 16$), 3 (I3; $n = 18$), and 4 (I4; $n = 20$). None of the children had received formal instruction in the Welsh language (vocabulary, spelling, or dictation) prior to the commencement of this study.

Materials

The pre- and post-test Welsh vocabulary items were taken directly from the school curriculum for that term. Two hundred Welsh vocabulary words were used that were derived from the following categories: School Uniform (34 words); Colours (20 words); Descriptors (54 words); School Mealtimes (64 words); and General (28 words).

These 200 words were further divided into three separate subtests that would be administered to children over three consecutive days (and would match the vocabulary content for the three SAFMEDS card packs that would be used during the intervention). Subtest one ($n = 67$: 34 School Uniform, 7 Colours, 26 Descriptors); subtest 2 ($n = 67$: 32 School Mealtimes, 7 Colours, 14 Descriptors, 14 General); subtest 3 ($n = 66$: 32 School Dinners, 6 Colours, 14 Descriptors, 14 General).

SAFMEDS had black text (Helvetica typeface, font size 16) printed on white card (9 x 4 centimetres) with each card having an English word on one side with the Welsh equivalent word on the other side.

Digital countdown clock timers were provided so that children could accurately time their sessions at the beginning of every class. The children graphed their data on Standard Celeration Charts (SCC). These charts are semi-logarithmic; they have a calendar x-axis and a logarithmic y-axis. This allows daily progress to be charted and proportional changes in performance to be graphically displayed (see for example Calkin, 2005).

Design

Because the children had already been pre-assigned to school classes, we used a quasi-experimental design. In addition, the school staff wanted the maximum number of children to receive the intervention, so that resulted in the study only having one WLC group. The design therefore had two factors: one between-group factor (Group: WLC vs. intervention groups I1, I2, I3, and I4) and one within-group factor (Test: pre- and post-test).

Procedure

Pre- and post-testing.

Pre-testing for all groups took place over three days during the first week of class, using one of the three subtests for each day. Children were instructed to write the Welsh word in the space next to the English word, not to worry if they did not know a word, but simply to move on to the next word. They were further instructed to bring their completed record sheets to the front of the class as soon as they had finished, so that children would not be able to rehearse their responses prior to the start of the study. The post-test used the same 200 words but was administered to children as one complete test: all the words were randomly sorted. Children followed the same procedure for post-tests as detailed above for the pre-tests.

Intervention

The week following pre-testing (week 2), each child in each of the intervention groups was given a pack of SAFMEDS at the start of each lesson, beginning with Pack 1. They also received data sheets and graphs to record and chart their daily data. The fluency aim for each pack was a frequency of ≥ 50 correct responses (hits) per minute with ≤ 2 errors (misses). If a child met fluency aims for three consecutive days, they would progress onto the next pack of SAFMEDS.

All children worked in pairs. At the beginning of each class the children would collect their work folder that contained all of their individual materials, and get their cards ready with the English writing facing toward them to prepare to carry out a warm-up practice session. Before they began, the teacher would remind them not to worry about knowing the words during the sessions, because this was not a test.

The first child of the pair would shuffle the pack to ensure that cards were in a random order and to work through the entire pack of SAFMEDS as quickly as possible. They were instructed to *see* the English word on the front of the card (read silently), and then to *say aloud* (speak) the Welsh equivalent of that English word. After they had said the Welsh word for the English word on the card they would flip the card over to check if they had said the correct Welsh word. If they were correct, they would move on to the next card. If they were not correct, they would say the Welsh word aloud. This meant that through participation in the practice session they would have said the Welsh word correctly once for every card in the pack.

The first child of the pair would then be ready to carry out a one-minute timing. They would set their countdown clock timer for 1 minute and get ready to begin their timing by shuffling the cards once again. They would then start the countdown clock timer—as with the warm-up procedure, they would read the English word from the front of the card silently and

speak the Welsh word aloud. They would then flip the card over for immediate feedback, either positive or corrective. Cards to which they had responded correctly (hits) were placed in a pile to their left hand side, whereas cards to which they had responded incorrectly (misses) were placed in a pile to their right hand side. They would continue to work through the cards following this procedure until the countdown clock timer rang (indicating the end of the 1 minute timing period) at which point they would stop. Following this, the children counted both the number of hits and misses and recorded each total separately on their Daily Scores Tables, and recorded their data on the graphs. On the chart a dot (•) was used for hits and an 'x' for misses. Once this was done, children would go through their misses again following the same procedure. This was repeated until they had made each card a hit (i.e., said one correct response to every card). At the end of this procedure, the second child from the pair would repeat the whole procedure—warm up and one-minute timing. The procedure took a maximum of five minutes for each child (10 minutes for the pair) at the start of the 50-minute class period. As children became more fluent on the content of the pack they were currently working with the time required decreased. The intervention was carried out three-times per week, and lasted for four weeks overall (weeks 2 to 5 of the study).

Results

A one-way ANCOVA was used to compare the post-test scores for the combined scores of the Intervention Groups ($n = 79$) with the Waiting List Control group ($n = 16$). The analysis revealed a significant effect for group, $F(1, 92) = 46.25, p < .001$, with the Intervention groups writing significantly more words correctly than the Waiting List Control group. Table 1 illustrates the associated means and standard deviations for the two groups. Using the Cochrane Method, an effect size was calculated using the *change scores* (post-test – pre-test) and SDs of the change scores for the Intervention and Waiting List Control groups (Higgins & Green, 2011, editors). The resulting effect size was large, $d = 1.88$, 95% CI [1.28-2.48]

INSERT TABLE 1

ABOUT HERE

Each intervention group's data were also analysed using a *paired sample t-test* to compare differences in performance between pre- and post-test for the number of Welsh words correctly written. It was found that all intervention groups pre- and post-test mean scores were statistically significant at least at a .001 significance level. Intervention Group 1 gained 36.96 words, $t(24) = 10.42$; Intervention Group 2 gained 37.47, $t(15) = 7.52$; Intervention Group 3 gained 20.89, $t(17) = 4.71$; Intervention Group 4 gained 52.87, $t(19) = 12.51$. These results clearly indicate that it was not any single intervention group that was driving the overall effect. Positive change was found in each of the four intervention classes.

Analysis of Individual Change

To enable the assessment of individual change, improvement scores were calculated for each child: the pre-test scores for each child were subtracted from their post-test scores. To ensure that our analysis was conservative the RCI was calculated using the means and standard deviations of the entire groups' scores at pre-test to calculate the SE. Multiplying the

SE by 1.96 provides a measure of magnitude of change required to be reliable at the $p < .05$ level (Evans, Margison, & Barkham, 1998; Jacobson & Truax, 1991; Zahra & Hedge, 2010). For calculation purposes, stability of the test scores were estimated by calculating the correlation for the scores in the control group between pre- and post-tests, as this would provide a more accurate measure of stability as the control group had not received the intervention.

We could further calculate the number of children from the Intervention Groups and Waiting List Control group for whom the intervention had been successful and represent the potency of the intervention using the Numbers Needed to Treat statistic (NNT; Altman, 1998; Barrowman, 2002; Bender, 2001; Lesaffre & Pledger, 1999; Pinson & Gray, 2003). The NNT provides us with a measure of the effectiveness of the intervention in terms of the number of the children required to receive the intervention to have one additional child achieve a positive outcome over and above the success rate in the comparison intervention. Therefore low NNT are preferable to high NNT.

Figure 1e highlights the results from the RCI analysis on the WLC group's data. It clearly shows that no child in this group met the criteria to achieve reliable change (increase of 23.07 words correct per minute or greater) when comparing their performance. Whilst nine children did improve, the other seven showed a decrease in performance. The poorest performing child achieved a deterioration of 24 words less than at pre-test (i.e., reliable deterioration).

Figure 1 (a-d) shows the performance of all children ($n = 79$) from the four intervention groups. It shows that 56 children (70.89%) in total achieved the RCI criteria, with only two of the remaining children showing a decrease in performance.

When considering the overall outcomes for all the children, the NNT is 2 (95% CI [1.2, 1.6]). This means that for every two children who received the intervention, one

additional child responded positively to the intervention than would have if they had received teaching as usual (i.e., what was provided for the WLC group).

INSERT FIGURE 1

ABOUT HERE

Discussion

The impetus and main aim for introducing this study into the school setting was that teachers had informed the researchers that children from these classes tended to do poorly on learning new Welsh vocabulary words, even though one of the main aims of the class was to achieve this very outcome.

The findings from this study illustrate support for this brief intervention that employed a direct method of teaching second-language vocabulary that was easy to implement, did not require specialised equipment, and was not costly in terms of time or other resources. The children in each of the intervention classes improved, showing that it was not just one of the intervention classes that was driving the effect. Rather, the intervention's effect was distributed through all the intervention classes. This is again highlighted in the large number of individual children who benefitted reliably from receiving the intervention.

A large group effect, and low NNT focused on individual child outcomes, were found when comparing the combined intervention groups with the control. Although there is no existing research providing comparison NNT figures for second-language vocabulary interventions, these low numbers do compare favourably to NNT rates for treatments for major depression (NNT between 3-5) and bulimia nervosa (NNT = 9) considered to be evidence-based (Pinson & Gray, 2003).

A secondary aim of the study had been to give responsibility to the teachers to implement and manage the study from the outset. This was designed to increase the likelihood that the methods could continue to be used in future classes without relying on support from the research team. Although we found that teachers successfully implemented the intervention in that outcomes for children were positive, we have no data on whether the teachers continued to use the intervention with other classes.

Despite the positive results, there are a number of limitations of the present research. The initial pre-tests were administered as three separate subtests over a consecutive three-day period (which tallied with the three SAFMEDS pack contents). However, the final post-test was given as one single test in just one sitting. This limitation in experimental control was due to the balance we were striving to achieve between running a research study in a real world setting whilst still enabling the teachers to be responsible for the implementation and management of the intervention within each classroom. We also did not carry out follow up testing of children at a period after post-tests. Such data would have been important to explore whether the gains for the intervention children continued to maintain over time in comparison to the Waiting List Control children. It would be also interesting in future research to conduct a more systematic analysis of errors that individual children made. Such error analysis could aid in individualising the instruction for children and hence lead to faster learning (Kubina & Yurich, 2012). A stronger research design would also have involved random allocation of individual children to the intervention and control groups.

Although all but two of the children made gains from baseline to post-test a number of children only made small gains. We have no additional data for the children in the intervention groups with regard to their attendance in class, the number of SAFMEDS packs they reached fluency aim in, or the total amount of practice they conducted throughout the intervention. Had these data been collected they might help explain why some of these children did not reach the criteria for reliable change. Future research should investigate the relationship between the gains and the amount of practice individual children conducted; that is, whether there is a ‘dose’ effect of using SAFMEDS.

From the results, we are not able to conclude whether the obtained effects then act as a pivotal language skill for better second language learning. As the children were all fluent English speakers we had them conduct a see English – say Welsh task. Our rationale here

was to prepare the children for speaking and using Welsh in the classroom and so we concentrated on the production of the Welsh words. It would be interesting to also see if the opposite relation would be as beneficial, that is, see Welsh – Say English. Future research should also attempt to better control in both consistency and standardisation of the measures used, and ensure that the outcome measures include tests of application of vocabulary in real world settings, such as in writing stories or conversational skills.

In the current study, we did not try to separate the effects of the SAFMEDS intervention from the process of charting and monitoring individual progress. Previous research has shown that teaching children to chart their learning can have motivational effects (Bower, 1985; Lindsley, 1995; Maloney, 1993). In future, researchers might attempt to pick apart these effects, perhaps by having three separate conditions: SAFMEDS only, SAFMEDS plus charting, and a Waiting List Control.

The teaching methodology employed for this study was inexpensive to implement, both in terms of time and resources. Initially, whilst children were learning the procedure the time taken was longer, but once children were conversant with the procedure it took approximately 5 minutes of lesson time at the start of each 50 minute class. The children were responsible for carrying out the entire procedure in their pairs, requiring only minimal teacher supervision, so it was not demanding on teacher time.

Anecdotally, children enjoyed this method of learning, as they would frequently already have begun working through their timings in pairs before the teacher had called the class to order to begin the classroom activities. This was also emphasised as certain children were requesting that they be allowed to take their cards home to practice (something that our initial research design would not allow). Again, anecdotally, the teachers reported an improvement in the punctuality of the children—if children were not there at the beginning of

the class they would miss both their timing and the opportunity to beat their previous personal best score.

As this research took place in a real world setting, and empowered the teachers to be the main instigators of the entire intervention, it lacked the control that could be achieved with a more rigorous experimental design. There is an opportunity in the future for this type of research to be conducted as a basic experiment and to address additional implementation questions. These might include the length of time it takes to become fluent on each set of SAFMEDS; how effective this fluent learning is in aiding recall after periods of time without practice, and what are the precise range of rates of responding that result in the proposed outcomes of fluency: Retention, Endurance, Stability, and Application (RESA; Binder, 1996). It is probable that it is only when these outcomes of fluency are apparent that vocabulary will be generalised into other second-language behaviours (e.g., using appropriate learned vocabulary in conversation, reading, and writing). Additionally, future research could be focused on the effectiveness of these techniques when applied to first-language vocabulary teaching.

References

- Altman, D. G. (1998). Confidence intervals for the number needed to treat. *BMJ*, 317(7168), 1309-1312.
- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English Language Learners. *Learning Disabilities Research & Practice*, 20(1), 50-57. doi:10.1111/j.1540-5826.2005.00120.x
- Barrowman, N. J. (2002). Missing the point (estimate)? Confidence intervals for the number needed to treat. *Journal of the Association of Computing Machinery*, 166(13), 1676-1677.
- Beaton, A., Gruneberg, M., & Ellis, N. (1995). Retention of foreign vocabulary learned using the keyword method: a ten-year follow-up. *Second language Research*, 11(2), 112-120. doi:10.1177/026765839501100203
- Bender, R. (2001). Calculating confidence intervals for the number needed to treat. *Controlled Clinical Trials*, 22(2), 102-110.
- Beverley, M., Hughes, J. C., & Hastings, R. P. (2009). What's the probability of that? Using SAFMEDS to increase undergraduate success with statistical concepts. *European Journal of Behavior Analysis*, 10(2), 235-247.
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, 19(2), 163-197.
- Binder, C., Haughton, E., & Bateman, B. (2002). Fluency: Achieving true mastery in the learning process. Retrieved from http://www.fluency.org/Binder_Haughton_Bateman.pdf
- Binder, C., Haughton, E. C., & Van Eyk, D. (1995). Increasing endurance by building fluency: Precision teaching attention span. *Journal of Precision Teaching*, 12(2), 29-34.
- Bloom, B. S. (1986). The hands and the feet of genius: Automaticity. *Educational Leadership*, 43(5), 70-77.
- Bower, R. (1985). Self-charting: Giving kids a chance. *Journal of Precision Teaching*, 6(1), 14-20.
- Calkin, A. B. (2005). Precision teaching: The standard celeration chart. *The Behavior Analyst Today*, 6(4), 207-213.
- Carter, R. (1987). Vocabulary and second/foreign language teaching. *Language Teaching*, 20, 3-16. doi:10.1017/S0261444800004225
- Claypool-Frey, R. (2009). SAFMEDS. Retrieved from <http://precisionteaching.pbworks.com/SAFMEDS>
- Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge, UK: Cambridge University Press.

- Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, 61(2), 367-413. doi:10.1111/j.1467-9922.2010.00613.x
- Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health*, 1(3), 70-72.
- Fredrick, L. D., & Hummel, J. H. (2004). Reviewing the outcomes and principles of effective instruction. In D. J. Moran & R. W. Malott (Eds.), *Evidence-based educational methods* (pp. 9-22). San Diego, California: Elsevier Academic Press.
- Georgiades, N. J., & Phillimore, L. (1975). The myth of the hero-innovator and alternative strategies for organisational change. In C. C. Kiernan & F. P. Woodford (Eds.), *Behaviour modification with the severely retarded* (pp. 313-319). North Holland: Elsevier Academic Press.
- Graf, S., & Lindsley, O. R. (2002). *Standard Celeration Charting*. Poland, Ohio: Graf Implements.
- Heward, W. L. (1994). Three low-tech strategies for increasing the frequency of active student response during group instruction. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman, & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 283-320). Pacific Grove, California: Brooks/Cole Publishing Company.
- Higgins, J. P. T., & Green, S. (2011, editors). *Cochrane handbook for systematic reviews of interventions*. Retrieved from <http://www.cochrane-handbook.org/>.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 21(02), 181-193.
- Hughes, J. C., Beverley, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis*, 8(2), 221-238.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19.
- Johnson, K. R., & Layng, T. V. J. (1996). On terms and procedures: Fluency. *The Behavior Analyst*, 19(2), 281-288.
- Johnson, K. R., & Street, E. M. (2004). *The Morningside model of generative instruction: What it means to leave no child behind*. Concord, MA: Cambridge Centre for Behavioral Studies.
- Kang, H.-W., & Golden, A. (1994). Vocabulary learning and instruction in a second or foreign language. *International Journal of Applied Linguistics*, 4(1), 57-77. doi:10.1111/j.1473-4192.1994.tb00055.x
- Kubina, R. M., & Yurich, K. K. L. (2012). *The Precision Teaching Book*. Lemont, PA: Greatness Achieved.

- Lesaffre, E., & Pledger, G. (1999). A note on the number needed to treat. *Controlled Clinical Trials*, 20(5), 439-447.
- Lindsley, O. R. (1995). Precision teaching: By teachers for children. *Journal of Precision Teaching*, 12(2), 9-17.
- Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning*, 24(6), 515-525. doi:10.1111/j.1365-2729.2008.00289.x
- Maloney, M. (1993). Teaching children to chart. *Journal of Precision Teaching*, 10(2), 18-19.
- Mueller Gathercole, V. C., Mon Thomas, E., & Hughes, E. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism*, 11(6), 678-720. doi:10.1080/13670050802149283
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. New York: Heinle-Cengage Learning.
- National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (Report of the National Reading Panel)*. Bethesda, MA: Author.
- Pinson, L., & Gray, G. E. (2003). Number Needed to Treat: Underused measure of treatment effect. *Psychiatric Services*, 54(2), 145-146.
- Roberts, W., & Hampton, E. (2008). Evaluating and sustaining Precision Teaching interventions in schools. *Debate*(127), 23-32.
- Roberts, W., & Norwich, B. (2010). Using Precision Teaching to enhance the word reading skills and self-concept of secondary students: a role for professional educational psychologists. *Education Psychology in Practice*, 26(3), 279-298.
- Vanderplank, R. (2009). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, 43, 1-37. doi:10.1017/s0261444809990267
- Vieitez, D. E. (2003, May). *Precision teaching and SAFMEDS in a college course*. Paper presented at the Association for Behavior Analysis 29th Annual Convention San Francisco.
- Wang, A. Y., & Thomas, M. H. (1995). Effects of keyword on long-term retention: Help or hindrance? *Journal of Educational Psychology*, 87(3), 468-475. doi:10.1037/0022-0663.87.3.468
- Zahra, D., & Hedge, C. (2010). The Reliable Change Index: Why isn't it more popular in psychology? *PsyPag Quarterly*(76), 14-20.

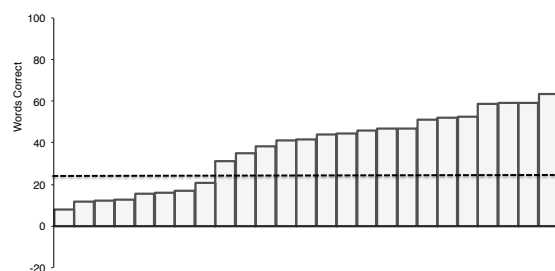
Table and Caption

Table 1

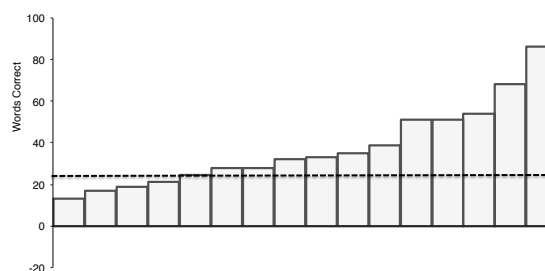
Means and Standard Deviations for the combined Intervention Groups compared to the Control Group.

	<u>Pre-test</u>	<u>Post-test</u>
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Waiting List Control	35.13 (20.12)	34.81 (22.58)
Intervention	37.03 (24.24)	74.57 (32.98)

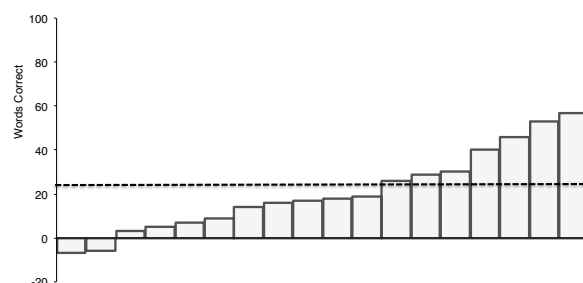
Figure and Caption



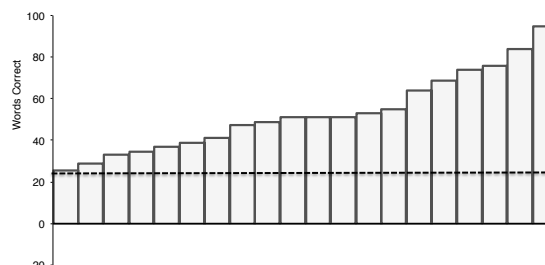
1a] Intervention Group 1 ($n = 25$); Reliable change achieved by 17 children



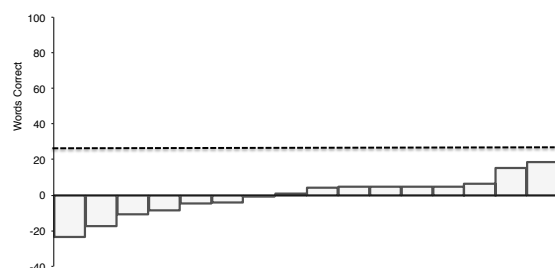
1b] Intervention Group 2 ($n = 16$); RCI achieved by 12 children



1c] Intervention Group 3 ($n = 18$); RCI achieved by 7 children



1d] Intervention Group 4 ($n = 20$); RCI achieved by 20 children



1e] Waiting List Control (WLC) Group ($n = 16$); RCI achieved by 0 children

Figure 1. Shows charts for each of the intervention groups and the WLC group respectively. Scores above zero are gains in performance from pre- to post-test in the number of Welsh words written correctly. Scores above the dotted line indicate that the improvement is reliable at the individual level as measured by the *Reliable Change Index* (RCI).